Harrison Eiteljorg II

Archiving Archeological Data in the Next Millennium

A few years ago a colleague and friend of mine called from Turkey to the institution where he had once taught. He was looking at materials from an excavation he had conducted while on the staff at that institution some 30 years earlier, and he needed information from his own records—records that remained in the archives of his former employer. Within a few days he had his answer.

My own archival materials are different from my colleague's; they are electronic files rather than paper records. As a result, were I to call from Athens to my own office on a similar mission, in this case to get some information from my computer-aided design (CAD) model of the older propylon, I fear that the outcome would not be as successful. Although the computer file could surely be located, I would need to offer instructions at every step of the process so that the file could be opened and the information obtained. That is, I would have to explain not only what information I need and where it might be, but what program to use and how to retrieve the specific information required. **How, then will someone else find that information years from now?**

My concern about the difficulty of retrieving information from my own archival repository prompts a broader concern for archival storage as we approach the new millennium. Archeologists desperately need access to data from past excavations, not just the publications (which too often do not appear) but the raw data collected in the field. If electronic data complicates the archival process, then **our apparent progress in record keeping may be regression instead.**

I intentionally painted a sharp contrast in the preceding. There are doubtless things in paper archives that are all but impossible to find, and there are surely things in electronic archives that are easy to find. Nonetheless, there are special problems with computer archives that should concern all archeologists, issues that affect the ease with which data can be retrieved and, more important, the security of data in an archive.

At this point I can turn this discussion in either of two directions: technical or practical. I can talk about the technical issues that will bedevil those of us interested in electronic archives, or I can talk about the more practical problems that will crop up as we try to use the technology and preserve the records that are its fruit. As an admitted technological optimist, I will assert that the technical issues can be solved; furthermore, I think they are of little or no interest to the majority of scholars. The practical problems, on the other hand, can only be solved by those archeologists who are prepared to do things that may be unappealing. So I think the real-world difficulties encountered by archeologists who must create and preserve electronic data are more interesting to readers—and more significant for the profession.

The practical problems relate to two different parts of the archeological process—first, general dig planning and direction and second, treatment of the electronic files at the end of the line. Let me start with the issues that surround general dig planning and direction.

General Dig Planning and Direction

Excavation directors are a bit like orchestra conductors; both direct specialists, each of whom must be able to do much of his/her work without the direct intervention of the leader. Both worry about choosing the specialists, how they all fit together, and timing. As the conductor does not tell the violinist how to tune his instrument, so the excavation director does not tell the pottery specialist how to construct a database for the pottery. The director will watch over the utility of the end product, not the details. That is true for all the specialists.

Over time, directors have learned that there are some unexpected things that must be watched, though those matters may seem to be beyond their ken. For instance, how many worried about the kind of paper used for notebooks in the 1930s or, in later periods, the permanence of the film stock used by staff photographers? As those turned out to be critical items, so issues surrounding software used on sites—specific programs chosen and modifications made—are now important issues that I do not think all directors have recognized and confronted. When software is chosen, for instance, directors must ask their specialists such pertinent questions as "In what forms will/can the data be stored?" or "Can I use this data on a MAC/PC since

you are going to store it on a PC/MAC?" or, most important of all, "Can we integrate your data with that of the other specialists?"

Some excavation directors are asking those questions now, but issues surrounding archival storage of the data are still missed too often. Most important are choices of computer file format. If data are stored in uncommon file formats, then, at the least, extra work will be required when the files are ultimately archived, because the format will have to be changed. In extreme cases, the data may be all but useless in electronic form, and it may be necessary to print everything on paper to preserve the information. I should point out that, for very long-term excavations, these questions of file formats can have more immediate repercussions. As a dig progresses, computer power surely will change, but software chosen may not. Therefore, consideration of file formats may be very important for the day-to-day operation of a dig. Should the chosen software cease to be the best for the work, it should—no, must—be possible to bring along the data as the underlying software is changed.

As I see it, then, excavations directors, as they plan and direct the recordkeeping process, should concern themselves with issues of data storage. They do not currently ask their architects what pens or papers they use, but they do ask about permanence of the drawings. The same concerns—with a few wrinkles due to the technology—apply to electronic data storage; so directors must make certain that the data files created are in useful, modern formats and can be moved, if necessary, to standard formats for archival storage or data transfer.

Preparing Files for the Archive

Now I want to turn to the problems with the treatment of electronic files at the conclusion of a project. I will assume that the data have been stored in appropriate formats. When the project is complete, the director is responsible for making the archival arrangements. I think the biggest problem here is that the director must take an active role in this process, though that may not have been required for archival preservation in the past. It may once have been possible to assume that archivists would, as they must, accept what they received (or retrieved from vacated offices) and make the best of it. Electronic files simply cannot be treated that way.

I will use my CAD model of the older propylon as an example here—a 3D computer model of the remains of the entrance structure for the Athenian Acropolis before 437 B.C. Leaving aside the potential problem of finding the file on its hard drive and recognizing it as worthy of archival treatment, an archivist must recognize the electronic file format and know whether it is current or not, whether there are more appropriate formats, whether there are standard formats for the particular data type, whether it must be transferred to one

Upcoming National Park Service Courses

If you are interested in either of the following two courses contact Gay Tracy, NEDCC, 100 Brickstone Square, Andover, MA 01810; Tel: 978-470-1010 or ccmail <tracy@nedcc.org>

Afterimages: Reformatting Visual Materials in a Digital World ($100 for a limited number of NPS staff, $275 for all others) will be held September 16th-18th, 1998, at the National Archives and Records Administration Archives II Facility in College Park, MD, and is co-sponsored by NPS, NARA, and NEDCC. **Afterimages** will teach archivists, curators, historic preservation specialists, librarians, and other CRM staff how to:

- plan and manage projects to reformat visual materials, such as cellulose nitrate and acetate negatives, including contracting with an outside vendor
- select and prepare visual collections for reformatting including preservation and handling
- understand best practices, benchmarks, and quality control for photographic and digital imaging
- manage legal issues, contracts, cost containment, and cost benefit analyses

School for Scanning: Issues of Preservation and Access for Paper-Based Collections ($255 before October 15, 1998, $325 thereafter) will be held December 7-9, 1998, at the Le Petit Theatre du Vieux Carre in New Orleans, LA, and is co-sponsored by the NPS, the Getty Information Institute, the Historic New Orleans Collection, SOLINET, and NEDCC.

This course will teach participants:

- the basics of digital technology, including deciphering digital jargon
- content selection for digitization including legal issues
- text and image scanning, including costs, quality control, metadata, and multi-versioning
- electronic publications

of those formats, and which format (if any) can be expected to remain current for a reasonable length of time.

Once the question of format has been dealt with, the archivist must confront questions relating to the use of the model by others. For instance, my model has more than 200 different data segments. In-situ stones of cut marble, with specific date span, and lying in the stair of the entrance structure are in one data segment; the nearby tripod base, also of cut marble and with the same dates, is in another. The particulars of these data segments are not important for this discussion, but the model cannot be used effectively without an understanding of the segments and the way they have been named. That information is not implicit in the model; it must be supplied in a set of documentation that I must have prepared.

Included in the model are blocks that I measured with tapes and line levels as well as blocks that I surveyed with photogrammetric techniques. As a result, there are different levels of confidence to be placed in different parts of the model. Users of the model need to know that so they can assess the accuracy of specific data. However, the difference between the parts of the model measured with tapes and the parts surveyed with photogrammetry are not apparent to a user of the model, and the difference cannot be determined with the model alone. Again, I must have supplied information about survey methods if the model is to be used to maximum advantage. There are also data files attached to the CAD model. The formats of the files, the fields used, the limits on terms used for the data, and much more must be given to users so that they can use these files as well.

Once more, I must have supplied that information along with the files themselves. I**n short, I, as supplier of the data, must have supplied considerable documentation along with the data files.** In an ideal world, there would have been similar documentation to accompany paper files, and that documentation has often been missing as well. However, there is a critical difference between the paper and electronic files. The paper files, by and large, can be used without the documentation. The terms can be teased out of internal relationships and usage; the organization can also be determined. It may take time, but it is possible. In the case of computer files, on the other hand, the relationships are often impossible to find, and the documentation is much more central to the utility of the files. Time simply may not be enough.

At the conclusion of a project, then, the director must produce the documentation required for archival storage—and he/she must do so quickly. The need for documentation has already been spelled out, but I believe that delay is an important problem as well. Not only is it easy to put off the work and, in the process, lose track of important information, the time lag also creates problems peculiar to electronic data. The longer the delay, the more likely it is that the files will be compromised by neglect. File formats may become obsolete, files may decay, or they may simply be lost. I do not believe they are safe when stored for long periods on institutional mainframes or servers; nor do I believe it is safe to leave them on an individual hard disk, not to mention a floppy disk. Time is of the essence.

Condusion

So the new millennium is coming. Along with it come new forms of data storage. We can't afford the old, casual processes for archiving in the new millennium with its new electronic forms of data. Individual scholars and institutions must examine their priorities, assign a higher level to archival storage, and insist on meeting the ethical requirements of archival preservation. The technical problems, I believe, are small. However, particularly if our past record—in terms both of publication and past archival preservation processes—is a guide to the future, our success is far from assured. I fear that in this work, as the cartoon character, Pogo, once said, "We have met the enemy and he is us."

Harrison Eiteljorg II is Director, Center for the Study of Architecture and the Archaeological Data Archive Project.

Assistant Chief Ranger Trischman of Yellowstone at Upper Lamar, Yellowstone. Anonymous.